An exploratory study of paper sharing in Mendeley's public groups

Huiqin Gao, Wuhan University
Changping Hu, Wuhan University
Tingting Jiang, Wuhan University

Abstract
Mendeley website is a representative academic social networking service. We aim to study how papers are shared in the public groups in Mendeley. The results show that 61.58% of the public groups were extremely small in size, containing only one member (the creator of the group). When it comes to paper sharing, 26.88% of the groups had no papers added to them. Large groups did exist, i.e. the groups having more than 1,170 members. Groups with large amount of papers also existed, i.e. groups having as many as 90,458 papers. On the other hand, there are top groups with high averages of paper readership; interestingly, these groups had small numbers of members and papers, both below 20. From the results of this research, the truth of online ecology on Mendeley website could be revealed. Taking an insight into the current condition helps group owners activate their groups, and also helps operators of Mendeley make decisions on improving services. Those improvements would make Mendeley a more advanced social platform for scholarly knowledge communication.
Keywords: online groups; academic social networking services (ASNS); altmetrics; virtual communities.
Citation: Gao, H., Hu, C., Jiang, T. (2015). An exploratory study of paper sharing in Mendeley's public groups. In *iConference 2015 Proceedings*.

Research Data: In case you want to publish research data please contact the editor.
Contact: gaohuiqin@foxmail.com, hcpwhu@163.com, tij@whu.edu.cn

1 Introduction

First launched in 2008, Mendeley[1] has developed into a useful tool for researchers and scholars to discover, organize, read, and cite academic papers. Each user has an academic profile and a personal library in the system. As a typical social networking service, Mendeley allows users to build social connections and form groups where research ideas are communicated. The goal of this study is to identify the influences of its public groups on paper sharing.

The design of Mendeley was inspired by Last.fm, the world's largest music sharing system, accommodating personalized recommendations, collaborative filtering, and ontological classifications of user generated resources(Henning and Reichelt, 2008). Referred to as "Academic Social Networking Services (ASNS)"(Oh and Jeng, 2011) or "Social Research Network Sites (SRNS)"(Bullinger et al., 2010) , Mendeley satisfies the academia's needs of social intercourse, supporting the self-presentation of one's academic background and achievements, the discovery of target researchers and scholars, as well as one-to-one communication through internal messages and group discussion. Similar systems include ResearchGate[2], Academia[3], and CiteULike[4].

There are three types of groups in Mendeley: private, upon-approval, and open groups. The latter two are public groups where the group details are completely visible to any users. But upon-approval groups require that users need to obtain the group owners' permission to join them. Group activities on Mendeley have been a popular research focus in the area of knowledge sharing and communication. A research team based on the University of Pittsburgh has probed into various topics concerning Mendeley groups, including the multi-disciplinary collaboration (Oh and Jeng, 2011), group owners' descriptions and group outcomes (Jeng et al., 2012), the interaction between disciplines (Jiang et al., 2013), and the participation of open group members (Jeng et al., 2014). In the most recent study (Jeng et al., 2014), they conducted a survey of Mendeley open groups and analyzed 146 responses. Based on the results they investigated users' common activities, usage habits, and motivations for joining groups.

[1] http://www.mendeley.com
[2] http://www.researchgate.com/
[3] https://www.academia.edu/
[4] http://www.citeulike.org/

These studies provide interesting insights into the open groups. However, they failed to consider upon-approval groups, and the influence of the papers shared in groups. The only mention of papers can be found in (Jeng et al., 2012) which examined the growth of group outcomes using a small sample of 529 groups. On the other hand, the behavior of group members have changed as Mendeley's user base kept evolving during the past two years. It is necessary to collect up-to-date data to reveal the status quo of public groups. Thus we aim to investigate paper sharing in Mendeley public groups. The following research questions are addressed in this our research:

RQ1: Are groups large in number of members? Are groups large in number of papers?
RQ2: Which disciplines have more groups?
RQ3: Are members active in adding papers?
RQ4: Are public groups valuable as accumulation of papers?

In this study, we used a Web crawler to capture group data on Mendeley. It is found that a considerable proportion of the groups are very small or inactive, i.e. containing only one member or no paper. These groups were not included in the analysis of group average readership. The average readerships of groups were calculated with readership statistics that is known as "altmetrics". A number of studies have been conducted on altmetrics using the readership data from Mendeley (Thelwall and Maflahi, 2014, Mohammadi and Thelwall, 2013, Bar-Ilan, 2014, Haustein and Larivière, 2014). The term of "altmetrics" was coined in 2010(Priem et al., 2010). It is used as an alternative indicator to paper citations (Haustein and Siebenlist, 2011) and for early estimation (Wang et al., 2014) to provide new insights to academic papers' impacts and usefulness via digital use and the analysis of sharing data(Kwok, 2013). The effectiveness and efficiency of altmetrics has been tested and verified by (Sud and Thelwall, 2014). Nowadays, altmetrics has been adopted by publishers including PLoS[5], Nature Publishing Group[6], and Elsevier[7]. Mendeley's readership statistics has being applied as a practical altmetrics measurement by researchers since 2012(Priem et al., 2012). The readership statistics is obtained by aggregating the behavior of many users(Henning and Reichelt, 2008). It referrs to the number of users who have collected a particular paper to their personal libraries in Mendeley[8].

Our research provides readers with an overall understanding of the public groups on Mendeley, enabling them to recognize the top groups with highest average readerships. On the other hand, group owners, active members, as well as Mendeley operators may find useful implications to increase the popularity of public groups and improve user participation.

2 Method

The data was collected from Mendeley. Although Mendeley provides API for researchers and developers, it fails to meet the integrated requirements of this research. Thus a web crawler in python was written for our data collection. The program conducted a top- down extraction, beginning with each discipline. For each group in a discipline, the following information was extracted and recorded: URL, name, discipline(s), number of members, and number of papers. For the papers shared in a group, the following information was extracted and recorded: title, first author's name, publication year, added date, and readership count. All the information was stored in a SQL database. The data collection started on 20 July 2014 and completed on 22 July 2014.

What deserves attention is the overlap between groups: 1) a member may appear in more than one group, thus the sum of members for all groups will larger than the real number of people involved; 2) a paper might also appear in more than one group, thus the sum of papers will also be larger than the real number of papers collected; 3) a group can be assigned by the creator to 1-3 disciplines as a cross-disciplinary group, and the sum of groups for 25 disciplines will be larger than the groups existed.

Mendeley counts readership for papers from the perspective of the whole website, as a result, the readership is not affected by the group overlap. For the cross- disciplinary groups, we used additional fields in the table to mark the other disciplines, namely "discipline_2" and "discipline_3". When analyzing

[5] http://www.plosone.org/static/almInfo.action
[6] http://www.nature.com/press_releases/article-metrics.html
[7] http://www.marketwatch.com/story/elsevier-announces-2012-journal-impact-factor-highlights-2013-07-15
[8] http://support.mendeley.com/customer/portal/articles/1626928-what-are-readership-statistics-

the characteristics for each group or each discipline, we used the intact data. When scanning the overall conditions, we used the distinct records, excluding the duplicated ones.

3 Results

3.1 Basic Characters

a) Members

A total of 106,156 distinct public groups were extracted. Compared to Wei Jeng's data of 34,508 groups in May 2012(Jeng et al., 2012), there was an increase during the past two years, indicating a flourishing vital force of Mendeley online groups. Table 1 shows the distribution of groups of different sizes.

Table 1. Number of Members in Public Groups

Members n. (mbr. n.)	Freq.	Percentage (%)
1	65,372	61.58%
2	20,388	19.21%
3-5	13,823	13.02%
6-10	3,865	3.64%
11-100	2,564	2.42%
101-500	132	0.12%
501-1170	12	0.01%
Sum	106,156	100%

Interestingly, most groups had only one member who was the creator of the group (N. = 65,372, 61.58%) or two members (N. = 20,388, 19.21%). The number of groups with three or more members is 20,396 (19.21%). The largest group had 1,170 members and its name was "Qualitative Research Methodology". This was a cross-disciplinary group under the disciplines of "Business Administration", "Management Science/Operations Research" and "Social Sciences".

b) Papers

There were a total of 5,034,736 papers shared in the 106,156 groups. It should be mentioned that they did not cover all the papers on Mendeley since the rest were not shared in public groups. They either did not belong to any groups or shared in private groups. Table 2 shows the number of papers in public groups, frequencies and percentages.

Table 2. Number of Papers in Public Groups

Papers n. (pp. n.)	Freq.	%
0	28,539	26.88%
1-10	34,369	32.38%
11-100	35,666	33.60%
101-500	6,273	5.91%
501-1000	694	0.65%
1001-5000	537	0.51%
5000-10000	55	0.05%
10001-15000	12	0.01%
15001-20000	5	0.005%
20001-90458	6	0.006%
Sum	106,156	100%

28,539 groups had no papers shared in them (26.88%), and 34,369 groups (32.38%) had only three to ten papers. The most frequent collection size is between 1 and 100, covering a percentage of 65.98%. Only 1.23% of the groups had more than 500 papers. The largest collection, containing 90,548 papers, appeared in the group named "Vaccine 2 " under the discipline of Biological Science. The number of groups having at least three members and at least one paper is 17,937, and we analyzed them further as "active groups". The other 88,219 groups are referred to as "inactive groups".

c) Papers per Member

The number of papers shared by each member partly reflects his/her activeness in group activities. For the 106,156 groups, we calculated the average papers added by members, as is shown in Table 3.

Table 3. Average Papers Added by Members in Public Groups

Papers per member	Freq.	Percentage
0	30,059	28.32%
1-5	30,562	28.79%
6-50	36,961	34.82%
51-500	7,927	7.47%
501-5000	620	0.58%
5001-45229	27	0.03%

For the groups with per capita papers below 0.5, it was counted as 0 in this part. Consequently the 30,059 in first row of Table 3 is higher than the 28,539 in first row of Table 2. The largest parts of groups had 6-50 papers per member, covering a percentage of 34.82%, followed by the groups with 1-5 papers per member, covering 28.79%. The highest value was 45,229 papers per member, and is of the group "Vaccine" as introduced in the former paragraph.

d) Overall

For both the total 106,156 groups and the active 17,937 groups, we calculated their minimum, maximum, mean and standard deviation, as is shown in Table 4.

Table 4. Descriptive statistics for 106,156 public groups

	Minimum		Maximum		Mean		Std. Deviation	
Groups corpora	**Total**	**Active**	**Total**	**Active**	**Total**	**Active**	**Total**	**Active**
Members n.	1	3	1,171	1,171	2.65	8.88	12.22	28.42
Papers n.	0	1	90,458	30,677	47.43	107.62	497.66	542.40
Papers per member	0	.0059	45,229	10,226	26.37	21.09	277.38	121.23

From Table 4 we can see that for the active 17,937 public groups on Mendeley, the average number of members is 8.88, average of papers is 107.62 and the average per capita added papers is 21.09. The maximum papers collection for the total 106,156 groups is 90,458 (group: "Vaccine"), larger than the maximum of active groups. This is because the group "Vaccine" has only two members and was excluded for later analysis.

3.2 Distribution in Disciplines

Table 5 shows the distribution of groups in disciplines, the top 5 and bottom 5 disciplines are covered.

Table 5. Discipline distribution of Mendeley public groups

Rank	Discipline	Percentage
1	Biological Sciences	19.45%
2	Computer & Information Science	17.68%
3	Medicine	10.55%
4	Social Sciences	6.32%
5	Engineering	6.29%
6	Education	6.06%
7	Psychology	4.18%
8	Business Administration	3.66%
9	Environmental Sciences	3.64%
10	Physics	2.73%
11	Chemistry	2.72%
12	Electrical and Electronic Engineering	2.34%
13	Economics	2.26%
14	Earth Sciences	2.10%
15	Humanities	1.83%
16	Arts and Literature	1.29%
17	Management Science / Operations Research	1.22%
18	Design	1.17%
19	Materials Science	1.03%
20	Astronomy / Astrophysics / Space Science	0.72%
21	Linguistics	0.72%
22	Mathematics	0.66%
23	Law; Sports and Recreation	0.49%; 0.49%
25	Philosophy	0.40%

Concerning the number of groups, the largest disciplines are Biological Science, Computer information science and Medicine. As was explored by Wei Jeng (2014), early users in Mendeley were mainly from disciplines of Computer & information science and Biomedicine. Thus it is not difficult to understand that these users created a large number of groups. Due to the work habits, researchers from disciplines of humanities including Linguistic, Law, and Philosophy do not engage a lot in computer-mediated work; as a result they do not create many public group on Mendeley.

a) Average Readership (AR)

In this article, the Average Readership (AR) of a group is defined as the average readership counts of all papers in this group. If a group has high AR, then the papers in this groups form a valuable accumulation of resources. If these groups could be well recognized and utilized, users of Mendeley would find their needed or interested papers more easily. AR is calculated as below:

$$AR\ (group)\ = \sum_{1}^{n} R/n$$

R: readership count of each paper in this group.
n: number of papers in this group.

Since Mendeley website was launched in 2008, so the readerships are all after year 2008, and the time span for calculating AR could only be from year 2008 to 2014.

b) High AR disciplines

We calculated the AR for the active 17,937 public groups and picked out the top 500 ones. Table 6 shows the top 5 disciplines with high AR groups.

Table 6. The top disciplines with groups of highest AR

Rank	Discipline	High AR groups n.
1	Computer & Information Science	132
2	Biological Sciences	118
3	Education	56
4	Medicine	29
5	Business Administration	26

The numbers of groups of Computer & Information Science, Biological Sciences and Medicine are in top 5 among the 25 disciplines, thus it is not strange to find they have more high AR groups.

Education has a collection size ranked 6th, and this discipline has the 3rd largest part of high AR groups. One possible reason is that many groups under this discipline are groups for courses in university, and in these groups many papers concerning learning or researching method were added, which receive high readership in the website. For example, the group with highest AR under discipline Education was named “Alfaisal English 1123”, and in this group there was a paper titled “How to choose a good scientific problem”, which had a readership count of 65,955, and is the highest in the whole website.

Business Administration has the 8th largest collection size, and the amount of high AR groups ranked the 5th.

c) High AR Groups

The top 20 group with highest Average Readership are shown as in Table 7.

Table 7. Top 20 groups with highest AR

Rank	Discipline	Name	mbr n.	pp n.	AR	AR/mbr	pp n./mbr
1	Engineering	Dottorandi, dottori e assegnisti di ricerca DISAT - Università dell'Aquila	5	3	52007	10401.4	0.6
2	Social Sciences	Curs administratius Mendeley	5	2	44003	8800.6	0.4
3	Biological Sciences	urban planning & ecology	6	51	33839	5639.83	8.5
4	Computer & Information Science	Medical Augmented Reality	4	2	32984	8246	0.5
5	Engineering	General Science/Philosophy	4	2	32983	8245.75	0.5
6	Computer & Information Science Philosophy Psychology	(Ex)-CogScis	7	4	32836	4690.86	0.57
7	Business Administration Management Science / Operations Research	Defence Acquisition Management Exec 6	12	4	32447	2703.92	0.33
8	Computer & Information Science Social Sciences	KAIST Library	10	7	24671	2467.1	0.7
9	Management Science / Operations Research	장지윤	3	13	22203	7401	4.33
10	Education	Alfaisal English 112	3	3	22010	7336.67	1
11	Computer & Information Science	CMP-G105	5	3	21995	4399	0.6
12	Medicine	NephroDD	3	3	21994	7331.33	1
13	Biological Sciences	GloNeuro@Mendeley	5	3	21992	4398.4	0.6

14	Medicine	Finance	8	23	20121	2515.13	2.88
15	Environmental Sciences	Direction de la Recherche Forestiere	10	5	19551	1955.1	0.5
16	Social Sciences	UQROO	10	11	17666	1766.6	1.1
17	Biological Sciences	Master 2 recherch GBCP	5	2	17340	3468	0.4
18	Biological Sciences	Grupo de Herpetologia	4	5	17226	4306.5	1.25
19	Chemistry	Chemists	3	4	16533	5511	1.33
20	Computer & Information Science	Competence modeling	5	4	16494	3298.8	0.8

As can be seen from Table. 7, the groups with high AR are commonly not large in size of members. The number of members range from 3 to 12, while the number of papers range from 2 to 51, and the ratio of papers per member range from 0.33 to 8.5, meaning that in these 20 highest AR groups, the average of papers per member are below 10. This phenomenon indicates that, there are no necessary connection between high average sharing ratio and high Average Readership.

Among the 20 top groups, 17 are assigned to one discipline, and the other 3 are cross-disciplinary groups. This indicates that cross-disciplinary groups do not have significant advantage of readerships.

4 Limitations

The data of Mendeley is changing in every second, and due to the time lag, some data might be missed during the operation of web crawler. Compared with the huge amount of total data, this minute missing could be accepted.

There might be some disorder of Mendeley's database and the operation of extraction program, as we discovered during research:

1) The total number of extracted public groups is 106,156, and 534 less than the published number of 106,690. This loss might be caused by the error of Mendeley database or the web crawler.

2) For some disciplines, the claimed number of groups did not equal the exhibited number, as Fig. II shows. Below the name of discipline, it says there are 20,164 groups, but besides the page numbers, it says there are 20,160 groups.

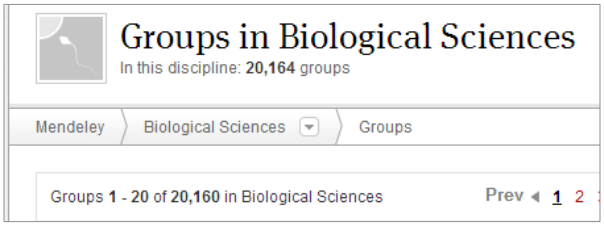

Figure 1. A mistake on Mendeley group page

3) In some groups, the list of papers showed repeated records. For example, in a group named "University of Kentucky, College of Public Health Faculty5 ", the content in "papers" column is totally the same on page from 48-55, under the "recent added" ranking order. This repeating brought disturbing duplicates to our database.

5 Discussion

5.1 Activation of the Groups

The high AR groups are generally not large groups, either small in size of members or in size of collections. This might provide guidance for group owners and the active members. As Wei Jeng had discovered (Wei Jeng, 2014), "altruistic motivation was one of the most critical reasons associated with

users' group engagement". Members are enthusiastic with sharing papers with others. According to the findings of this research, influence is not reflected with the quantity of the papers shared. Many of the papers do not receive significantly higher readership after being added to groups. Nevertheless, it is better sharing papers with higher values. Besides the activities of members, there are supposed to be mechanism improvements from the perspective of Mendeley, as discussed below.

5.2 Inspiration for Service Mechanisms of Mendeley

a) Recommendation for Groups

There are many nearly dead groups in Mendeley, having very few members and papers. Moreover, the Average Readership of these groups are low. When a user searches for groups on a specific topic, all groups with names matching the topic will be listed. The results cannot be sorted by the number of members or papers or the Average Readership. This can be considered by Mendeley. There are recommendations for large groups, or rich groups, or high AR groups on the navigating page of disciplines.

b) Recommendation for Papers

Recommendation are provided on the navigating page of each discipline, recommending the most popular papers in this discipline. But there is no recommendation within groups. For each group, users are unable to rank the papers by readership count, only can they rank by alphabetic order or by date. If Mendeley enables ranking by readership, users can find popular papers more efficiently.

c) Merging of Groups

As we found from the dataset, many groups had the same names or very similar names. Furthermore, the disciplines are usually also the same. This means that the groups are dispersing users and valuable resources. This is time-consuming when users are using groups to discover papers. Thus, we recommend that there should be group merging mechanisms on Mendeley that combine groups with the same or similar names with the owners' permissions.

d) Extra Supports for Users from Humanities

The number of groups and value of AR were low under the disciplines of humanities, including Philosophy, Law, Linguistic, etc. This situation may be a result of the work habit and low participation in computer-mediated working processes. However, the online papers are still important resources for their scientific research, and the researchers would be benefiting, if they were utilizing the resources to the greatest extent. Thus there are should be extra support on Mendeley, to facilitate these users who are less familiar with computer operation or Internet utilization.

6 Conclusion

Mendeley badges "Mendeley Advisors" among users, to encourage them to "spread the word" of Mendeley at local level, as a global strategy of marketing, to increase the population of new users. However, it more important to keep the current users and enhance their usage of the services. And the proper way to achieve this is to improve the quality of services rather than depending on marketing strategies.

In-depth and comprehensive insights into the ecology of virtual communities on Mendeley help build objective cognition of the status quo of the services. This research collected global newest data and analyzed upon it. On this basis, a substantial part of the groups were found to be inactive. Among the active ones, descriptive statistics were calculated, and the disciplines were reviewed. The findings and discussion make contributions to the activation of Mendeley online ecology and directions for improving Mendeley's services.

Further research can follow in two directions. First, it is worthwhile to find out the reasons why many groups with a couple of members have large collection sizes and high Average Readership. For example the group "Vaccine" had only two members but 90,458 papers and an Average Readership as high as 483. It would be laboursome to add an average of 45,229 papers to the group, for either of the two users. Second, one can explore why so many groups under the discipline of Business administration have high ARs.

References

Bar-Ilan, J. 2014. Astrophysics publications on arXiv, Scopus and Mendeley: a case study. *Scientometrics,* 100**,** 217-225.

Bullinger, A. C., Hallerstede, S. H., Renken, U., Soeldner, J.-H. & Moeslein, K. M. 2010. Towards research collaboration–a taxonomy of social research network sites.

Haustein, S. & Larivière, V. 2014. Mendeley as a Source of Readership by Students and Postdocs? Evaluating Article Usage by Academic Status.

Haustein, S. & Siebenlist, T. 2011. Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics,* 5**,** 446-457.

Henning, V. & Reichelt, J. Mendeley - A Last.fm for research? 4th IEEE International Conference on eScience, eScience 2008, December 7, 2008 - December 12, 2008, 2008 Indianapolis, IN, United states. Inst. of Elec. and Elec. Eng. Computer Society, 327-328.

Jeng, W., He, D. & Jiang, J. 2014. User Participation in an Academic Social Networking Service: A Survey of Open Group Users on Mendeley. *arXiv preprint arXiv:1401.6495*.

Jeng, W., He, D., Jiang, J. & Zhang, Y. 2012. Groups in Mendeley: Owners' descriptions and group outcomes. *Proceedings of the American Society for Information Science and Technology,* 49**,** 1-4.

Jiang, J., Ni, C., He, D. & Jeng, W. Mendeley group as a new source of interdisciplinarity study: How do disciplines interact on mendeley? 13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2013, July 22, 2013 - July 26, 2013, 2013 Indianapolis, IN, United states. Institute of Electrical and Electronics Engineers Inc., 135-138.

Kwok, R. 2013. Research impact: Altmetrics make their mark. *Nature,* 500**,** 491-493.

Mohammadi, E. & Thelwall, M. Assessing the mendeley readership of social sciences and humanities research. 14th International Society of Scientometrics and Informetrics Conference, ISSI 2013, July 15, 2013 - July 20, 2013, 2013 Vienna, Austria. International Society for Scientometrics and Informetrics, 200-214.

Oh, J. S. & Jeng, W. Groups in Academic Social Networking Services--An Exploration of Their Potential as a Platform for Multi-disciplinary Collaboration. Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), 2011. IEEE, 545-548.

Priem, J., Piwowar, H. A. & Hemminger, B. M. 2012. Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv preprint arXiv:1203.4745*.

Priem, J., Taraborelli, D., Groth, P. & Neylon, C. 2010. *Altmetrics: A manifesto*.

Sud, P. & Thelwall, M. 2014. Evaluating altmetrics. *Scientometrics,* 98**,** 1131-1143.

Thelwall, M. & Maflahi, N. 2014. Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology***,** n/a-n/a.

Wang, X. W., Mao, W. L., Xu, S. M. & Zhang, C. B. 2014. Usage history of scientific literature: Nature metrics and metrics of Nature publications. *Scientometrics,* 98**,** 1923-1933.

Table of Figures

Table of Tables